

Experimental Designs

Alan Baron

University of Wisconsin-Milwaukee

In his preface to *Tactics of Scientific Research (Tactics)*, Murray Sidman voices some doubts about the wisdom of writing a book about experimental methods.

The pursuit of science is an intensely personal affair. Experimenters cannot always tell us how or why they do what they do, and the fact that their conclusions are sound so much of the time remains a puzzle even to the many philosophers, logicians, and scientists who have devoted a major portion of their time and effort to this problem. I do not claim to be a systematizer or even a classifier of the rules of experimental practice. Nor do I claim to be a spokesman for any unified group. Even those who find their activities most accurately described here would feel uncomfortably restricted if they had to proceed solely as I have outlined. Neither the practice of experimentation nor evaluation of its products can be bounded by any specific rules—a qualification that lends a certain note of irony to any book on experimental methodology. (Sidman, 1960, p. vi-vii)

Despite this disclaimer, there can be little question that Sidman, writing in 1960, was the spokesman for a small band of researchers with strong convictions about how research should be conducted. Their views were not popular among the larger group of experimental psychologists of that time, something that may help explain the forcefulness of Sidman's presentation. His goal is to explain and justify the methods of "the experimental analysis of behavior"—research that focuses on the steady-state behavior of individual organisms. But in the course of doing so, he engages in stinging criticism of the alternative and then dominant method of experimental psychology—so-called group-statistical research designs.

As Blough (1961) pointed out when he reviewed *Tactics*, Sidman goes well beyond simply drawing a distinction be-

tween the two methods (so far that "he will leave many of his readers behind"). Other writers since then have advocated the conciliatory position that both methods are needed on the grounds that each has its peculiar limitations as well as strengths (e.g., Kazdin, 1982), but Sidman does not take this easy path. To the contrary, he is uncompromising in pressing the argument that the data generated by the two methods are not only different but incommensurable (cf. Krantz, 1971). The difference is so fundamental that the data "represent, in a very real sense, two different subject matters." Indeed, two kinds of science are involved. Researchers must take a stand "as to which of these types of data, individual or group, will form the basis of the science they are trying to build" (Sidman, 1960, p. 54).

The criticisms that Sidman leveled against group-statistical designs were compelling 30 years ago and they remain so. In its most basic version, the traditional experiment compares groups of subjects who differ in their exposure to the experimental treatment. To handle uncontrolled differences from subject to subject, assignment to conditions is random, and enough subjects are included within each group to average out deviant cases. Inferential statistics then are used to determine whether between-group variation (differences between the group averages) is sufficiently larger than within-group variation (differences among subjects within the same condition) to warrant the conclusion that the treatment had a reliable effect.

Although the limitations of this strategy had been alluded to earlier (Skinner, 1938), Sidman's *Tactics* provides the detailed refutation. The average performance of a group cannot be counted on to properly represent the individual members. The problem is compounded when averages from different groups are used to depict functional relationships

The views expressed here emerged from discussions with students and colleagues, in particular, Mark Galizio, Michael Perone, Cindy Vance, and Marshall Dermer. Their help is gratefully acknowledged. Reprints may be obtained from the author, Department of Psychology, University of Wisconsin-Milwaukee, Milwaukee WI 53201.

because a function based on between-group data has no counterpart in the behavior of the individual organism. The most insidious aspect of group-statistical methods is that they lead the researcher away from an experimental analysis. Experimental control is replaced by statistical control, and the researcher must settle for procedures that test the reliability of experimental effects against "chance" (which is, quoting Boring, no more than "a synonym for ignorance," Sidman, 1960, p. 45).

Single-subject designs, by comparison, do not have these problems. The inquiry focuses directly on the ultimate concern of the research—the behavior of the individual organism—and a small number of subjects is studied at length rather than a large number for brief durations. Extraneous variables are controlled within the experiment rather than averaged out statistically. Effects of variables across their range of influence—functional relationships—are examined as they naturally occur, within the same organism rather than as a construction from group performances. The need for inferential statistics is obviated because behavior is observed as a steady state.

To judge from the research literature, Sidman's methodological precepts were taken quite seriously by the behavior-analytic researchers of his time. This may easily be seen by scanning the early volumes of the *Journal of the Experimental Analysis of Behavior (JEAB)*, the journal in which Sidman and his colleagues published. In the 1959 and 1960 volumes, for example, the vast majority of experiments were designed along within-subject rather than between-group lines (close to 90% of the articles according to my count). Most experiments included no more than 5 or 6 subjects; detailed information was provided about individual performances, usually through cumulative records (75%), and inferential statistics rarely were mentioned (less than 5%). A striking indication of the extent to which these experiments diverged from the customary practices of that time may be found in Krantz's (1971) analysis of the "separate worlds of operant and non-

operant psychology." He compared the 1969 issues of *JEAB* with those of the *Journal of Comparative and Physiological Psychology (JCPP)*, a journal that also published research on questions of animal learning. On the fundamental matter of whether the experiment pursued a functional analysis, Krantz found that within-subject designs, although common in *JEAB* (88%), were quite infrequent in *JCPP* (7%).

Sidman's *Tactics* and *JEAB* are, of course, closely linked. When *Tactics* was published, *JEAB*, begun three years earlier, was just coming into its own as an outlet for the methods advocated by Sidman. In his study, Krantz (1971) interviewed key researchers, many of whom held editorial positions during the early years of the Journal. They told him that *JEAB* was founded because "most operant conditioners, in submitting their papers to the *Journal of Experimental Psychology* and the *Journal of Comparative Psychology*, found a non-receptive editorial policy centered around the employment of single subjects measured across time" (p. 63). The same theme appeared in the reminiscences of *JEAB*'s founders on the 30-year anniversary of the Journal. A recollection by Thom Verhave is particularly appropriate. "I remember tales of woe by Murray Sidman about hostile letters by the Editor of the *Journal of Comparative and Physiological Psychology*. ANOVA [analysis of variance] had become the dominant mode of doing research and single-subject research as once pursued by Donders, Fechner, Ebbinghaus, Thorndike, Pavlov, and Yerkes was no longer fashionable and acceptable" (Verhave, 1987, p. 463).

Not only is Sidman's *Tactics* the seminal work on the methods of the experimental analysis of behavior (the "Bible," so to speak), but the book remains the primary source whenever the procedures and concepts of the single-subject method are brought up. After Skinner, Sidman is the most frequently cited author by Johnston and Pennypacker (1980) in their more recent treatment of methodology. Barlow and Hersen's (1984) book on

"single case experimental designs" refers to *Tactics* as "the definitive methodological treatise" (p. 30). And those research methods texts that now include a chapter on single-subject designs (not all do—and some place the method under the heading, "quasi-experimental") attribute the approach to Sidman's *Tactics*.

But looking back through the filter of 30 years, one begins to discern a second irony that I doubt was anticipated by Sidman. Given the force of his arguments for a single-subject approach (plus the success with which it was used by the behavior-analytic researchers of that time), it would not have been overly optimistic to expect that by now the method would be the dominant tactic of scientific psychology. This has not yet come to pass, however, and it surely is a disappointment that efforts to convert researchers of the traditional persuasion have not been more successful. But matters may be much worse than that. An impression I have held over the years is that more and more researchers *within* the behavior-analytic camp are becoming practitioners of the very methods deplored by Sidman. My worst fears were confirmed when I recently put this belief to the test. Close to one-third of the experiments in the 1989 issues of *JEAB* now are using one version or another of the inferential statistical procedures so harshly condemned in *Tactics* (up from 5% in 1959–60). Between-group designs also have risen, from 10% earlier to 30% at present, and the demise of the cumulative record so poignantly noted by Skinner (1975) is quite apparent (no more than 10% of the articles included them). The words of Pogo, the comic strip character of the 1960s, may be appropriate here. "We have met the enemy and they are us."

What is one to make of this? The unsuspecting reader might suppose that during the 30-year interim the arguments of *Tactics* had been refuted or in some way found inadequate. If this is so, the refutation is not one that has appeared in print. In fact, I am hard-pressed to cite publications in which the writers take serious issue with Sidman's criticisms of inferential statistics and between-group

designs. Moreover, objections to single-subject designs (many of which are mentioned and countered in *Tactics*) are not all that frequent. When they do appear, the theme is that single-subject designs are inappropriate for certain questions not ordinarily considered within the domain of experimental psychology (such as surveys of public opinion or evaluations of the effectiveness of community interventions). Another possibility is that the experimental analysis of behavior has become increasingly concerned with issues not anticipated by Sidman, those for which statistical-group designs are better suited (perhaps experiments with human subjects or experiments on processes that are deemed irreversible). But I could not find evidence for this either in current issues of *JEAB*. Although the subject matter of behavior-analytic research undoubtedly has changed over the years, many of the experiments which used group-statistical procedures could have proceeded along the lines recommended in *Tactics* had the researchers chosen to do so.

I am left with the interpretation that these evolving research practices reflect a growing conviction that the two designs are, for most purposes, interchangeable. Certainly, this is the message from the other side in that the journals that previously rejected single-subject experiments now publish them side by side with experiments cast in the group-statistical mold. Perhaps it is only right that a similar tolerance should grace the pages of *JEAB*. The disturbing feature of this way of resolving the tension is that it has come about without real discussion of the potential costs. Failure to recognize the difference "can lead to a hopeless confusion of basically incompatible data and principles" (Sidman, 1960, p. 54). Yet, current readers of *JEAB* (as well as of the traditional journals) are regularly placed in the position of having to interchange results and conclusions from the opposing methods.

The purpose of my comments is to bring to light discrepancies between the rules set forth in *Tactics* and the ways in which experiments sometimes are con-

ducted. On this and other questions, recognition of the difference between precept and practice can only be a healthy step. The resolution, however, is something else again. One starting point might be the clarification of the different roles played by inferential statistics in single-subject and between-group designs. More often than not, statistical analyses of results from single-subject experiments appear a last-ditch effort to make sense of less-than-orderly results. It is difficult to defend statistical remedies for poor experimental control (the practice was severely criticized in *Tactics*, as well as more recently, e.g., Michael, 1974), particularly when the data are collected within the laboratory rather than applied settings. Nevertheless, a number of the *JEAB* experiments which used inferential statistics fall in this category.

The conceptual issues are different for between-group experiments, if only because the statistical comparisons are more required than optional. The justification for between-subject comparisons is either that the variable under study has irreversible effects (e.g., severe punishment may permanently suppress responding), or that the variable, by its nature, is not subject to experimental manipulation (e.g., the subject's species). Under these circumstances, improved experimental control undoubtedly can benefit the analysis, in that reduced subject-to-subject variation within the groups makes the average difference between the groups more apparent. Nevertheless, one is left in the end with the need for rules to help determine whether the difference is reliable, that is, the rules provided by inferential statistics. (Single-subject methods for comparing steady states of the same subject, discussed in *Tactics*, are of no help here because the comparison involves different subjects). It can be seen that the researcher concerned with variables that cannot easily be manipulated faces an unhappy choice. He or she either must adopt practices that violate the requirements of a functional analysis or must abandon research on interesting and important topics. For the reasons given in *Tactics*, one would not question the desirability of being able to identify

the controlling relationship within the same subject. What is the researcher to do when this simply is not possible?

The dilemma is particularly apparent when operant conditioning is studied with human subjects (Baron & Perone, 1982). Humans bring a range of behavioral differences with them to the laboratory. Some can be controlled within the confines of the experiment, as when extended baselines allow pre-existing reactions to extinguish. Others, however, such as differences related to age or gender, defy ready solutions and, as a consequence, demand research designs that take between-subject differences into account (old vs. young; males vs. females). These "individual differences" can hardly be ignored if they are critical for the behavioral processes under study. Moreover, age or gender may be of interest in their own right.

Consider, for example, the potential contribution of the subject's age to the results of operant experiments. Regardless of the question under study, the researcher who ignores this characteristic does so at his or her peril because age is so strongly related to temporal aspects of performance (response rates as well as reaction times). Not quite as obvious is that a comprehensive account of the conditioning process requires the analysis of developmental variables. The issues here are not unlike those that pertain to phylogenetic differences. It has come to be recognized that an organism's phylogenetic status can place constraints on conditioning. Developmental status has similar potential for interactions with the conditioning process (Baron, Myerson, & Hale, 1988). No doubt, for researchers with patience and time there is the possibility of examining short-term developmental changes in humans on a within-subject (longitudinal) basis. As a practical matter, however, the study of human development calls for cross-sectional designs—comparisons of subjects of differing ages—in which case the usual statistical considerations apply.

I end by noting that there are a few stirrings in other quarters about the issues raised here. Perone (in press), for example, has discussed between-subject

comparisons as they relate to developmental questions, and Johnston (1988) recently considered the appropriateness of such designs in the evaluation of behavior modification procedures. These discussions focused on research questions that may require deviations from a functional analysis (for the reasons given above). The more difficult issue that must be broached concerns the justifications for deviations when a functional analysis otherwise is possible.

REFERENCES

- Barlow, D. H., & Hersen, M. (1984). *Single case experimental design: Strategies for studying behavior change* (2nd ed.). New York: Pergamon Press.
- Baron, A., & Perone, M. (1982). The place of the human subject in the operant laboratory. *The Behavior Analyst*, 5, 143-148.
- Baron, A., Myerson, J., & Hale, S. (1988). An integrated analysis of the structure and function of behavior: Aging and the cost of divided attention. In G. Davey & C. Cullen (Eds.), *Human operant conditioning and behavior modification* (pp. 139-166). New York: Wiley.
- Blough, D. S. (1961). Review of "Tactics of scientific research: Evaluating experimental data in psychology". *Contemporary Psychology*, 6, 278-280.
- Johnston, J. M. (1988). Strategic and tactical limitations of comparison studies. *The Behavior Analyst*, 11, 1-9.
- Johnston, J. M., & Pennypacker, H. S. (1980). *Strategies and tactics of human behavioral research*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kazdin, A. E. (1982). *Single-case research designs. Methods for clinical and applied settings*. New York: Oxford University Press.
- Krantz, D. L. (1971). The separate worlds of operant and non-operant psychology. *Journal of Applied Behavior Analysis*, 4, 61-70.
- Michael, J. (1974). Statistical inference for individual organism research: Mixed blessing or curse? *Journal of Applied Behavior Analysis*, 7, 647-653.
- Perone, M. (in press). Single-subject designs and developmental psychology. In H. W. Reese (Ed.), *Life-span developmental psychology: Methodological innovations*. New York: Academic Press.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.
- Skinner, B. F. (1938). *Behavior of organisms*. New York: Appleton-Century.
- Skinner, B. F. (1975). Farewell, my LOVELY! *Journal of the Experimental Analysis of Behavior*, 25, 218.
- Verhave, T. (1987). Reminiscences of JEAB. Faded images. *Journal of the Experimental Analysis of Behavior*, 48, 462-464.